INFORMATION AND CONTROL 5, 108–143 (1962)

The Design of Nonlinear Filters and Control Systems. Part I*

JACOB KATZENELSON AND LEONARD A. GOULD

Electronic Systems Laboratory, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

The objectives of this paper are to provide a systematic analytical approach to the synthesis of continuous nonlinear filters and to apply the results to a variety of problems, e.g., filtering of signals from noise, characterization of nonlinear systems, and the design of compensation networks for control systems. The problems are illustrated in Fig. 1.

The work starts from the concept of a functional as the mathematical representative of a system. The optimum possible functional for any particular problem is approximated by a finite number of Volterra kernels. A typical question which the paper attempts to answer is:

"Given an input (signal plus noise) which is a sample function from a stationary, ergodic, random process and using the mean-square error criterion, what is the optimum filter consisting of a finite number of Volterra kernels to filter the signal from the noise?"[1]

This type of question leads to a set of simultaneous integral equations for which an iterative method of solution is provided. Examples and applications to filtering and control are discussed.

Part II of the paper will describe an experimental application of the theory to the characterization of the servomechanism associated with the pupil of the human eye. A measure of the complexity of the experiment will be developed and the applicability of the method to real problems will be discussed from this point of view.

INTRODUCTION

While simplicity is the obvious advantage in designing a linear system a question which usually remains unanswered is the amount, in

* This work was made possible by the support extended M.I.T. by the U. S. Army Research Office under Contract No. DA-ARO(D)-31-124-G193 and was carried out as part of a doctoral program in Electrical Engineering.

[1] A filter which consists of $h_0, h_1, h_2, h_3, \cdots, h_n$ is called a filter of degree n and sometimes will be denoted as $\langle h_0, h_1, h_2, \cdots, h_n \rangle$. Every kernel h_j is called a kernel of order j. Its corresponding filter is a filter of order j.

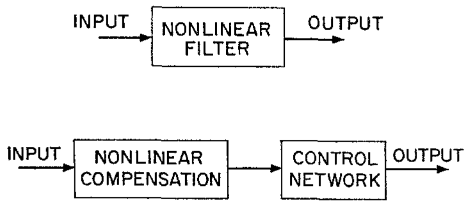

FIG. 1. Nonlinear filtering and nonlinear (cascade) compensation

terms of system performance, which one pays for the simplifying assumption. Although this question seems to haunt the field of linear systems from its early beginning, until recently, very little systematic work has been done in the area of nonlinear systems.

In nonlinear systems, as in linear systems, the first problem which one has to face is how to specify and characterize a nonlinear time invariant system. In answering this question the system engineer is looking for a convenient, workable description of the output in terms of the input.

One possible way to describe a nonlinear system is by the differential equation which relates the input and the output (Cunningham, 1958; Minorsky, 1957). This approach has the same disadvantage which the linear differential equation has in linear theory; it is not convenient from the system point of view (Bagdadi, 1961). Usually, the differential equation is an implicit description of the input–output relation and the equation by itself does not specify the system where by the term "system" we mean the assignment of an output to every input. Additional conditions have to be given to assure uniqueness, for example, initial conditions in a transient problem.

A systematic approach to the characterization of nonlinear systems by an explicit description of the input–output relation was started by Prof. Norbert Wiener. His work, which is summarized in (Wiener, 1958), was followed by the works of Singleton (1950), Zadeh (1953), Bose (1956), Barrett, Brilliant (1958), Zames (1961), and others. (See Zadeh (1961) for a survey of the area.)

The main new concept established by these works is the view of the functional as the mathematical equivalent of a system. A function associates a value $f(x)$ with each value of the independent variable within some domain of the independent variable. A functional connects a value, $F(x(t))$, to every function $x(t)$ which is in some domain of definition in function space. Similarly, a system connects a value, the present output, to a function, the past input. This point of view reduces the problem of characterization of a system with a given class of inputs to characterization of a functional which operates on a certain class of functions. In these works the class of input functions is an ergodic, stationary, random process and the criterion of performance is the mean-square error.

The various works in this field can be classified according to the way they answer two basic questions; First, what assumptions are made about the input random process, and second, what form is imposed on the optimum filter?

The work of Wiener (1958) and Barrett assume a complete statistical knowledge of the input—a gaussian process. In this case it is convenient to use the special properties of the gaussian process and describe the system in terms of orthogonal functionals. These are the Hermite-Laguerre polynomials and convolution-like integrals operating on Hermite polynomials.

A knowledge of second order statistics $p(x(t_1), x(t_2))$ which have a diagonal expansion (class Λ) is assumed by Barrett and Lampard (1955). It was shown by Zadeh (1957) that filters of a certain class are given by very simple relations when the input signal is of class Λ.

Other works which concern special input processes are those by Nuttall (1958), George (1959), and Chesler (1960).

All the above work is dependent upon special properties of the inputs. The primary advantage is that the result is expressible in a neat analytical form.

Works by Bose (1956) and Zames (1961) describe methods for a general type of input. Bose's gate functions remain orthogonal for any kind of input process. Therefore, this method is suitable for an orthogonal expansion of the filter regardless of the properties of the random process.

Zames (1961) characterizes the system and the probability distribution of the past of the signal by a Wiener expansion and provides a way of constructing Wiener filters for nongaussian noise.

The method which is described in the present work provides a systematic analytical approach to the design of a continuous filter. The input can be any type of random process provided certain averages exist. The work starts from the description of a continuous functional by Volterra integrals (Volterra, 1930).

$$y(t) = h_0 + \int_\Omega h_1(\tau)x(t - \tau)\, d\tau + \int_\Omega\int_\Omega h(\tau_1, \tau_2)x(t - \tau_1)x(t - \tau_2)\, d\tau_1\, d\tau_2 + \int_\Omega\int_\Omega\int_\Omega h(\tau_1, \tau_2, \tau_3)x(t - \tau_1)x(t - \tau_2)x(t - \tau_3)\, d\tau_1\, d\tau_2\, d\tau_3 + \cdots$$

where Ω is the domain of integration (the domain on which $x(t)$ is defined.) Frechet (1910) proved that any continuous functional $y[x(t)]$ defined on $\{x(t)\}$ with $\{x(t)\}$ a set of continuous functions defined on a finite interval $[a, b]$ can be represented by Volterra integrals. Brilliant (1958) reproved the theorem, extended the proof to an infinite interval and thoroughly investigated its meaning for system applications. It appears as if this representation was first used for circuit analysis by Wiener in 1942.

Generally speaking we would like to answer a question of the following nature: Assume that the best possible filter (Doob, 1953) (a filter which performs the best operation for filtering a signal from noise, for example, under some general restriction like physical realizability.) can be represented by this series. How can we find the h_n, $n = 1, \cdots, \infty$, which specify the filter? At present the answer to this question is not known except in some very special cases concerning gaussian processes. On the other hand it is doubtful whether we really need the expression for the optimum possible filter for any type of random process. In general, we can expect this filter to be given by a considerable number of high-order kernels. From the engineering point of view the construction of higher— and higher— order kernels might be quite complicated and, in many cases, impractical.

Instead of an expression for the optimum possible filter we would like to have a method of approximation which starts with simple elements and, by increasing the complexity, is able to approximate the optimum possible filter to an arbitrary degree of accuracy. The last point is important as it assures that the method by itself would not impose limitations on the performance of the filter.

Of course, any approximation is expected to yield better performance than Wiener's optimum linear filter, with the cost, naturally, of increasing the complexity of the system. These problems motivate the following question which is typical of the questions which this work attempts to answer:

"Given an input (signal plus noise) which is a sample function of a stationary random process and using the mean-square error criterion, what is the optimum filter of degree n (a filter which consists of $h_1, h_2, \cdots, h_n$) to filter the signal from the noise?"[1]

A similar question is asked and answered in this work with regard to prediction, filtering, and control compensation networks for nonminimum phase linear networks and linear networks with constraints (for linear compensation see Newton et al., 1957). The filter which is defined by the above question is one possible approximation to the best possible filter which is expressed by an infinite number of kernels. Obviously, taking the optimum filter of degree n is better than or equivalent to taking the first n terms in the Volterra series of the best possible filter. With the increase of the number of kernels taken, the error $\overline{\epsilon_N^2}$ decreases. As n goes to infinity the filter with a finite number of kernels approaches the optimum possible filter. When viewed from this point of view Wiener's optimum linear filter (Wiener, 1949; Lee, 1960) is the case $n = 1$. It is clear that by taking higher-order kernels the performance is improved.

It is also clear that this method provides a way to start from relatively simple elements and increase the complexity as the accuracy requirements are increased. Here, of course, a question might arise as to the meaning of simplicity and complexity. For example, when h_2 is considered, some second-order terms, like combinations of linear networks and squarers, are quite simple to realize whereas others which may involve multipliers are complicated. As a result one can expect that, after the kernels are found, some additional approximations will be made. However, no further restriction on the class of functionals can be made without giving up the requirement that the approach be both systematic and uniform.

Although the expression for the optimum possible filter is not very important, the knowledge of $\overline{\epsilon_\infty^2}$, the least possible error, as compared with $\overline{\epsilon_n^2}$, is of value, as this will determine the accuracy of the approximation and therefore the number of kernels, and perhaps which kernels one will choose to build. So far, it is not known how to calculate $\overline{\epsilon_\infty^2}$, and it is necessary to evaluate the kernels h_n in order to find $\overline{\epsilon_n^2}$.

Our work describes the problem, determines the integral equations which define the solution in terms of the h_i, and provides an iteration method for finding these h_i's. The advantage of this iteration method is that at each step one has a physically meaningful result which is a filter that is optimum in some sense. From the second step on the result is better than that of the linear optimum filter. Every additional iteration step decreases the mean-square error. It is proved that the iteration process converges to the optimum filter.

The number of the kernels h_i, $i = 1 \cdots n$, which are chosen at the beginning of the design procedure is arbitrary, but the calculation of the mean square error, $\overline{\epsilon_n^2}$, after each iteration step and the checking of the improvement achieved by adding an additional kernel h_{n+1} indicates how profitable it is to increase the accuracy of the solution or the number of kernels involved.

As will be shown later, the statistical data needed to specify a filter of degree n are the first $2n$ autocorrelation functions and the first n cross-correlation functions. This indicates a systematic approach to the characterization of an arbitrary random process which is motivated by the type of operation (in this case, filtering) which one intends to perform on the process, namely, characterization of a process by its various correlation functions.

To illustrate this point let us assume that one has to characterize noise at the output of some system. The way to get the statistical properties of the noise is to measure statistical averages. The mean and standard deviations will do if a very limited knowledge of the noise is required. A measurement of the autocorrelation function and cross-correlation with some desired signal would be the next step. This will provide all the information necessary for the design of a linear filter. If it is assumed that the foregoing measurements contain all the possible statistical information about the noise, it is in fact assumed that the noise is gaussian and thus the kind of filter one must construct is determined. If more information is sought about the noise, higher-order autocorrelations will be measured. The random process is, in effect, defined in terms of these autocorrelations and, accordingly, more complex nonlinear filters can legitimately be constructed.

NONLINEAR CORRECTIONS TO AN OPTIMUM LINEAR FILTER

This section treats a problem which is different from the design of the optimum nth-order filter. Although different, the problems are related and the solution of the simpler problem which is presented here gives

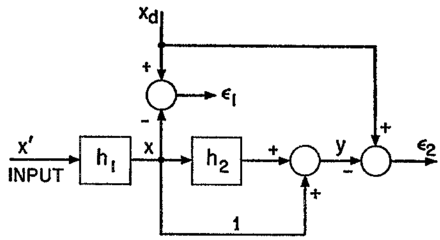

FIG. 2. A cascade configuration of nonlinear filters

some insight to the filtering question and shows the way for solving the more general problem.

The problem is illustrated by Fig. 2. In Fig. 2, x' is a sample function from a stationary ergodic random process; x_d is the desired output—a stationary ergodic random process related to x'; h_1 is the optimum physically realizable linear filter which minimizes the mean-square difference between its output, x, and x_d, $\overline{(x_d - x)^2}$.[2] In cascade with h_1 there is a second-order filter, h_2, in parallel with an identity operator, 1. The output of the whole filter is denoted by y. Clearly, h_2 is a second-order correction to the optimum linear filter. The question to be answered is: What is the optimum second-order kernel h_2 which minimizes $\overline{(x_d(t) - y(t))^2}$?

$$\overline{(x_d - y)^2} = E\left\{(x_d(t) - x(t) - \int_0^\infty \int_0^\infty h_2(\alpha, \beta) x(t-\alpha) x(t-\beta)\, d\alpha\, d\beta)^2\right\} \tag{1}$$

Now, $x_d(t) - x(t)$ is actually the error which remains after filtering the signal with an optimum linear filter. Let us denote the error at the output of the linear filter by ϵ_1 and the error at the output of the whole filter by ϵ_2. Using this notation in Eq (1) we get

$$\overline{\epsilon_2^2} = \overline{(\epsilon_1(t) - \int_0^\infty \int_0^\infty h_2(\alpha, \beta) x(t-\alpha) x(t-\beta)\, d\alpha\, d\beta)^2} \tag{2}$$

The problem is to find $h_2(\alpha, \beta)$ which minimizes the above expression. In order to find the condition which h_2 has to fulfill to minimize $\overline{\epsilon_2^2}$ one uses the calculus of variations in the same way as it is used to

[2] $E[x]$ is the mathematical expectation of x and is also represented by $\bar{x}$.

obtain the Wiener-Hopf equation for the optimum linear filter. (Wiener, 1949; Levinson, 1949; Lee, 1960). The result is:

$$\phi_{\epsilon_1 xx}(\tau_1, \tau_2) = \int_0^\infty \int_0^\infty \phi_{xxxx}(\tau_1 - \tau_2, \tau_1 - \alpha, \tau_2 - \beta) h_2(\alpha, \beta)\, d\alpha\, d\beta \qquad \text{for } \tau_1 \geqq 0, \tau_2 \geqq 0 \tag{3}$$

where

$$\phi_{\epsilon_1 xx}(t_1, t_2) = \overline{\epsilon_1(t)x(t - t_1)x(t - t_2)} \tag{4}$$

and

$$\phi_{xxxx}(t_1 - t_2, t_1 - \alpha, t_2 - \beta) = \overline{x(t - \alpha)x(t - \beta)x(t - t_1)x(t - t_2)} \tag{5}$$

Note that Eq. (3) is a two-dimensional linear integral equation of the first kind and gives the unknown kernel h_2 in terms of two correlation functions—a third-order correlation, $\phi_{\epsilon_1 xx}$, between the error, ϵ_1, and the signal, x, and a fourth-order autocorrelation function, ϕ_{xxxx}, of the signal, x. This is directly analogous to the Wiener-Hopf equation for an optimum linear filter:

$$\phi_{xx_d}(\tau) = \int_0^\infty \phi_{xx}(\tau - \alpha) h(\alpha)\, d\alpha \qquad \text{for } \tau \geqq 0 \tag{6}$$

However, there is a major difference between the two equations. While the integral on the right-hand side of Eq. (6) is a convolution-type integral, the right-hand side of Eq. (3) is not a convolution in two variables since ϕ_{xxxx} depends on $\tau_1 - \tau_2$ in addition to $\tau_1 - \alpha$ and $\tau_2 - \beta$. Consequently, the Wiener-Hopf technique (spectrum factorization) is not generally suitable for solving this equation. Only in the case where the fourth-order correlation function of the signal is factorable in the form

$$\phi_{xxxx}(t_1 - t_2, t_1 - \alpha, t_2 - \beta) = \hat{\phi}(t_1 - t_2)\phi(t_1 - \alpha, t_2 - \beta) \tag{7}$$

does Eq. (3) reduce to a Wiener-Hopf type integral equation and a spectrum factorization technique can be then used to solve the problem. Only the general case will be considered here. The special case of Eq. (7) and its solution by a multidimensional spectrum factorization will be treated in a future paper.

Returning to Fig. 2 it is clear that an equation similar to Eq. (3) would result if we preferred to correct h_1 by a third-order filter, h_3, a

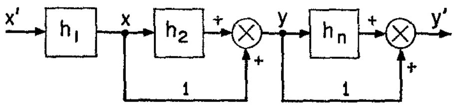

FIG. 3. Example of cascade configuration

fourth-order filter, h_4, or a filter of order n. Similarly, instead of having h_1 as the first filter in Fig. 2, one can have any type of filter there and correct it with an arbitrary h_n. For example, the place of h_1 in Fig. 2 can be taken by h_1 with an optimum second-order correction. This filter might be improved by adding an nth-order correction giving rise to the configuration of Fig. 3.

Thus the general problem of this section is, in fact, the designing of an optimum correction of degree n to an existing filter in the configuration of Fig. 2. When the existing filter is an optimum filter of its kind, the solution of the problem will provide a method of improving the filter by adding a cascade correction. The resulting equations are similar to the Wiener-Hopf equation for the linear filter where the place of x_d is taken by the error which remains after the filtering by the first stage.

The following two questions given some insight into the cascade configuration:

1. Using the configuration of Fig. 3, where h_1 is the optimum linear filter, can the performance of h_1 be improved by putting a linear filter, k_1, in place of h_2? Obviously, the combination of h_1 and k_1 is still a linear filter and as h_1 is the optimum linear filter, it cannot be improved by k_1. However, it is of interest to see how this question is answered by considering the integral equation which corresponds to Eq. (3).

The integral equation for k_1 is

$$\overline{\epsilon_1(t)x(t-\tau_1)} = \int_0^\infty k_1(\alpha)\, \overline{x(t-\alpha)x(t-\tau_1)}\, d\alpha \qquad \text{for } \tau_1 \geqq 0 \quad (8)$$

The error at time t, $\epsilon(t)$, and the signal at time t_0, $x(t_0)$, where $t_0 < t$, are uncorrelated (Davenport and Root, 1958). As the error has mean zero, the left-hand side of (8) is zero for $\tau_1 > 0$.

$$0 = \int_0^\infty k_1(\alpha)\overline{x(t-\alpha)x(t-\tau_1)}\, d\alpha \qquad \tau_1 \geqq 0$$

A filter which fulfills Eq. (8) would not have any effect on the mean square error as can be concluded from

$$\overline{\epsilon_2{}^2} = \overline{\epsilon_2{}^2} - 2\int_0^\infty \overline{k_1(\alpha)\epsilon_1(t)x(t-\alpha)\,d\alpha}$$
$$+\int_0^\infty\int_0^\infty k_1(\alpha)k_1(\beta)\overline{x(t-\alpha)x(t-\beta)}\,d\alpha\,d\beta$$

Therefore $k_1(\alpha) = 0$ will do just as well.

2. Let both x' and x_d be gaussian processes. Can the performance of h_1 be improved by a correction of degree n? The question will lead to an equation which is similar to Eq. (3):

$$\overline{\epsilon_1(t)x(t-\tau_1)\cdots x(t-\tau_n)} = \int_0^\infty\cdots\int_0^\infty h_n(\alpha_1,\cdots,\alpha_n)$$
$$\cdot\overline{x(t-\alpha_1)\cdots x(t-\alpha_n)x(t-\tau_1)\cdots x(t-\tau_n)}\,d\alpha_1\cdots d\alpha_n \tag{9}$$
$$\text{for } \tau_1\cdots\tau_n > 0$$

Since in gaussian processes no correlation between the error and the signal implies independence,

$$\overline{\epsilon_1(t)x(t-\tau_1)\cdots x(t-\tau_n)} = \overline{\epsilon_1(t)}\,\overline{x(t-\tau_1)\cdots x(t-\tau_n)}$$

As the mean of $\epsilon_1(t)$ is zero, the left-hand side of Eq. (9) is zero and, again, $h_n = 0$ will do just as well. One obtains a variant of the well known fact that the best possible filtering of a gaussian process can be done by a linear filter.

In the following we shall consider the solution of Eq. (3) as the method of its solution is the same as that of an equation involving a kernel of any order n.

Let us note the following properties of the equation and the random process. We assume that the even-order autocorrelation functions are bounded. Autocorrelations of order 2^n are bounded by their value at the origin. Even-order autocorrelation can be bounded by suitable moments. This can be deduced by successive use of the Schwartz inequality. In the same way and from the assumption that $\overline{x_d{}^2}$ is bounded it follows that the crosscorrelations of the type $\phi_{\epsilon xx}(\alpha, \beta)$ are also bounded.

$\phi_{xxx}(\tau_1 - \tau_2, \tau, - \alpha, \tau_2 - \beta)$ is completely symmetrical in τ_1, τ_2, α, and β. In particular, $\phi_{xxxx}(\tau_1, \tau_2, \alpha, \beta) = \phi_{xxxx}(\alpha, \beta, \tau_1, \tau_2)$.

Equation (3) can be written in the form:

$$f(t) = \int_\Omega K(t, s)h(s)\,ds \qquad \text{for } t > 0 \tag{10}$$

where s denotes the combination of variables α, β and t denotes the combination t_1, t_2; $f(t)$ denotes $\phi_{\epsilon_1 xx}(t_1, t_2)$ and $K(t, s)$ is $\phi_{xxxx}(t_1, t_2, \alpha, \beta)$. The domain of integration is denoted by Ω. We can eliminate the condition $t \geqq 0$ by multiplying the two sides of the equation by $1(t_1) \times 1(t_2)$, where $1(t)$ denotes a step function $[1(t) = 1, t > 0;\ 1(t) = 0, t < 0]$. If we again denote $1(t_1) \times 1(t_2) \times \phi_{\epsilon_1 xx}(t_1, t_2)\ \phi_{\epsilon_1 xx}(t_1, t_2)$ by $f(t)$ and $\phi_{xxxx}(t_1, t_2, \alpha, \beta)1(t_1)1(t_2)$ by $K(t, s)$, Eq. (10) becomes

$$f(t) = \int_\Omega K(t, s)h(s)\, ds \tag{11}$$

As before, K is symmetrical and bounded. A property of K required in the sequel is that $K(t, s)$ is square integrable. This means that

$$\int_\Omega \int_\Omega K^2(s, t)\, ds\, dt < \infty$$

If one is designing filters which operate on a finite part of the past, Ω is a finite interval and the integrable square property of K follows from the fact that it is bounded.

For a filter which operates on the infinite past, this is not necessarily so. In order that $K(s, t)$ be square integrable, $K(s, t)$ as a function of s, t has to attenuate "fast enough" as the variables s and t approach infinity so that the integral $\int\int K^2(s, t) ds\, dt$ will converge. However $\phi_{xxxx}(t_1 - t_2, t_1 - \alpha, t_2 - \beta)$ is not L^2 (square integrable) in t_1, t_2, α, β, as its value remains constant. This difficulty can be overcome by attenuating the past of the signal. We assume that at time t the filter h_n operates on $x(t - \tau)e^{-k\tau} (k > 0)$ instead of operating on $x(t - \tau)$. This means that the past is attenuated exponentially according to its distance from the present. Using the attenuated past Eq. (3) becomes:

$$\begin{aligned} \phi_{\epsilon_1 xx}(\tau_1, \tau_2)e^{-k(\tau_1+\tau_2)} = \int_0^\infty \int_0^\infty & h(\alpha, \beta)\phi_{xxxx}(\tau_1 - \tau_2, \tau_1 - \alpha, \tau_2 - \alpha) \\ & \cdot e^{-k(\tau_1+\tau_2+\alpha+\beta)}\, d\alpha\, d\beta \qquad \text{for } \tau_1 \geqq 0, \tau_2 \geqq 0 \end{aligned} \tag{12}$$

Again let

$$f(t) = \phi_{\epsilon_1 xx}(\tau_1, \tau_2)e^{-k(\tau_1+\tau_2)}1(\tau_1)1(\tau_2)$$

$$K(s, t) = \phi_{xxxx}(\tau_1 - \tau_2, \tau_1, - \alpha, \tau_2 - \alpha)e^{-k(\alpha+\beta+\tau_1+\tau_2)}1(\tau_1)1(\tau_2)$$

$$h(s) = h_2(\alpha, \beta)$$

Then we get

$$f(t) = \int_\Omega K(s, t)h(s)\, ds \tag{13}$$

The kernel K is still symmetric and in addition it is now L^2. The same operation converts $f(t)$ into an L^2 function. It has to be emphasized that with respect to the mean square error nothing is lost by introducing the weighing function. In fact, one can view the attenuation in the following way: Let $y(t)$ be the output of a second order filter h_2

$$\begin{aligned} y(t) &= \int_0^\infty \int_0^\infty h_2(\alpha, \beta)x(t - \alpha)x(t - \beta)\, d\alpha\, d\beta \\ &= \int_0^\infty \int_0^\infty h_2(\alpha, \beta)e^{k(\alpha+\beta)}x(t - \alpha)e^{-k\alpha}x(t - \beta)e^{-k\beta}\, d\alpha\, d\beta \end{aligned} \tag{14}$$

If

$$k_2(\alpha, \beta) = h(\alpha, \beta)e^{+k(\alpha+\beta)}$$

then

$$y(t) = \int_0^\infty \int_0^\infty k_2(\alpha, \beta)x(t - \alpha)e^{-k\alpha}x(t - \beta)e^{-k\beta}\, d\alpha\, d\beta \tag{15}$$

Comparison of Eq. (14) with Eq. (15) shows that the use of a weighing function does not change the output, the input, or the filter, but changes what is called the signal and what is called the filter.

From the fact that $K(s, t)$ is a bounded symmetric function it follows (Courant and Hilbert, 1953; Smithies, 1958) that $K(s, t)$ possesses a set of eigenfunctions Ψ_i and eigen values λ_i such that

$$\lambda_i \int_\Omega K(s, t)\Psi_i(s)\, ds = \Psi_i(t)$$

$f(t)$ is an L^2 function, and, as shown in Appendix A, it can be specified in the mean by the eigenfunctions of $K(s, t)$.

Therefore, the operator

$$h_n(t) = \sum_{i=1}^{n} \lambda_i(f, \psi_i)\Psi_i(t) \tag{16}$$

where

$$(f, \Psi_i) = \int_\Omega f(t)\Psi_i(t)\, dt$$

solves Eq. (13) in the mean (Appendix A). That is,

$$f(t) = \underset{n\to\infty}{\text{l.i.m.}} \int_\Omega h_n(s)K(s, t)\, ds \tag{17}$$

Uniqueness of the solution can be demonstrated in a manner similar to that used for the general case treated in a subsequent section (see Appendix B). $h_n(t) = \sum_1^n \lambda_i(f, \Psi_i)\Psi_i(t)$ can be considered to be an approximate solution of the integral equation and an approximation to $h(t) = \lim_{n\to\infty} h_n$. However, h certainly cannot be considered a function from a strict point of view. Nothing has been said in any sense about convergence of the series $\sum_{i=1}^{\infty}\lambda_i(f, \Psi_i)\Psi_i(t)$. Indeed, according to Picard (Smithies, 1958), a necessary and sufficient condition for h to be an L^2 solution of the equation is

$$\sum_{i=1}^{\infty} \lambda_i^2(f, \Psi_i)^2 < \infty$$

So far, it is not clear what conditions the process has to satisfy in order that the above condition will be fulfilled.

However, the properties of h as a function are of no interest to us, as the kind of solution which we seek for the integral equation is an operator and not a function. As a representation of an operation (or generalized function) the series $\text{l.i.m.}_{n\to\infty} \sum_1^n \lambda_i(f, \Psi_i)\Psi_i(t)$ has to converge after operating on a function of a suitable class, that is, an input which is a sample function of the given random process. (For generalized functions and the concept of convergence of a generalized function see Kolmogorov and Fomin (1957).) Unless the problem is of such a nature that the eigenfunction of $K(s, t)$ can be calculated easily, and the series

$$\sum_{i=1}^{\infty} \lambda_i(f, \psi_i)\psi_i(t)$$

can be summed to a net analytical expression, filters involving an infinite number of eigenfunctions would not be built. Therefore, it is of interest to investigate the properties of the approximation

$$h_n(t) = \sum_{i=1}^{n} \lambda_i(f, \psi_i)\psi_i(t)$$

where only a finite number of eigenfunctions are used to construct the filter. One can start from the assumption that $K(s, t)$ is given and has to be approximated by n normalized orthogonal functions $\Psi_i(t)$, $\phi_i(t)$

so as to minimize

$$\int_\Omega \int_\Omega (K(s,t) - A_n(s,t))^2 \, ds \, dt$$

where

$$A_n(s,t) = \sum_{i=1}^{n} \frac{\Psi_i(t)\phi_i(s)}{i}$$

It is found (Courant and Hilbert, 1953) that the best approximation to $K(s, t)$ consists of the first n eigenvalues and corresponding eigenfunctions of the series $\lambda_1, \lambda_2, \cdots$ where $|\lambda_i| \leqq |\lambda_{i+1}|$ (all i).[3] After this approximation has been made the optimum filter is given by a finite number of terms:

$$h(t) = \sum_{i=1}^{n} \lambda_i(f, \psi_i)\psi_i(t)$$

However, the problem is not the approximation of $K(s, t)$ but the minimization of the mean-square error. The question to be answered is: What n eigenfunctions should be chosen in order to minimize the error? For any n eigenfunctions chosen, the error is (see sequel)

$$\overline{\epsilon_2^2} = \overline{\epsilon_1^2} - \sum_{i=1}^{n} \lambda_i(f, \psi_i)^2 \tag{18}$$

Minimization of the expression involves both $f(t)$ and $K(s, t)$ and cannot be done without more information about $f(t)$.

Another question of some interest is: What are the best n normalized orthogonal functions to minimize $\overline{\epsilon^2}$? This question leads to complicated equations and so far it has not been answered.

The Mean-Square Error

When the filter $h_n(t)$ is used, the mean-square error is

$$\begin{aligned} \overline{\epsilon_2^2} &= \overline{\epsilon_1^2} - \int_\Omega h_n(t)f(t)\,dt + \int_\Omega\int_\Omega K(t,s)h_n(t)h_n(s)\,ds\,dt \\ &= \overline{\epsilon_1^2} - \int_\Omega h_n(t)f(t)\,dt \end{aligned} \tag{19}$$

$$\overline{\epsilon_2^2} = \overline{\epsilon_1^2} - \sum_{i=1}^{n} \lambda_i(f, \Psi_i)^2 \tag{20}$$

[3] The process for finding eigenvalues which is described by Courant and Hilbert (1953) yields the eigenvalues arranged in the above order.

When the number of terms taken approaches infinity, the error becomes

$$\epsilon_2^2 = \overline{\epsilon_1^2} - \int_\Omega h(t)f(t)\,dt \tag{21}$$

$$= \overline{\epsilon_1^2} - \sum_{i=1}^{\infty} \lambda_i (f, \Psi_i)^2 \tag{22}$$

When written in full, Eq. (21) becomes

$$\overline{\epsilon_2^2} = \overline{\epsilon_1^2} - \int_0^\infty \int_0^\infty h(\alpha, \beta)\, \overline{\epsilon_1(t) x(t-\alpha) x(t-\beta)}\, d\alpha\, d\beta \tag{23}$$

This equation is similar in form to the equation for the error which remains after filtering with an optimum linear filter (Wiener, 1949)

$$\overline{\epsilon_1^2} = \overline{x_d^2(t)} - \int_0^\infty h(\alpha) \overline{x_d(t) x(t-\alpha)}\, d\alpha$$

where $x(t)$ is the signal and $x_d(t)$ the desired output. It is interesting to note again that, in the design of h_2, $\epsilon_1(t)$ appears in the place where $x_d(t)$ appears in the design of a linear filter.

OPTIMUM FILTERS WITH A FINITE SET OF KERNELS

In this section the methods which were developed above are used to solve the general problem: Given an input x and a desired output x_d which are sample functions of a stationary ergodic random process; what is the filter of degree n which minimizes the mean-square error

$$\overline{\epsilon^2} = \overline{(x_d - y_n\{x(t)\})^2},$$

where $y_n\{x(t)\}$ denotes the output of the filter?

A filter which answers the above question has the canonical configuration of Fig. 4, where a filter of the second degree is described. Consider the series configuration of Fig. 2 with h_1 as the optimum linear filter and

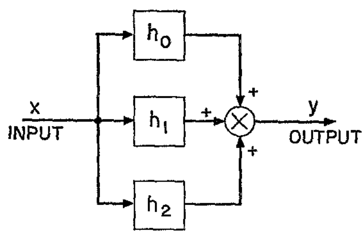

FIG. 4. A second order filter; parallel configuration

h_2 as an optimum second-order correction. This filter could not be the best possible filter of the second degree or the best possible filter of degree n if we continue further the cascade, as we do not change the first stage, h_1, when a second stage is added. There is no reason to assume a priori that the linear filter of a second-degree optimum filter is equal to the optimal linear filter which operates on the same process.

The method of solving the above problem is demonstrated by using a filter of the second degree. This particular example is used since the representation of the solution by using a general filter of degree n complicates the equations without giving more information or insight than the chosen example.

In Fig. 4 let h_0 be a zero-order filter (output is equal to h_0 regardless of the input), and let h_1, h_2 be first- and second-order physical realizable filters, respectively. The mean-square error at the output is

$$\overline{\epsilon^2} = E\left\{\left(x_\mathrm{d} - h_0 - \int_0^\infty h_1(\alpha)x(t-\alpha)\,d\alpha - \int_0^\infty\int_0^\infty h_2(\alpha,\beta)x(t-\alpha)x(t-\beta)\,d\alpha\,d\beta\right)^2\right\} \tag{24}$$

The kernels h_0, h_1, h_2 which minimize the above expression are given by the following equations:

$$h_0 = \overline{x_\mathrm{d}} - \int_0^\infty h_1(\alpha)\overline{x(t-\alpha)}\,d\alpha - \int_0^\infty\int_0^\infty h_2(\alpha,\beta)\overline{x(t-\alpha)x(t-\beta)}\,d\alpha\,d\beta \tag{25}$$

$$\begin{aligned}\int_0^\infty h_1(\alpha)\overline{x(t-\alpha)x(t-\tau_1)}\,d\alpha &= \overline{x_\mathrm{d}(t)x(t-\tau_1)} - h_0\,\overline{x(t-\tau_1)} \\ &\quad - \int_0^\infty\int_0^\infty h(\alpha,\beta)\overline{x(t-\alpha)x(t-\beta)x(t-\tau_1)}\,d\alpha\,d\beta \\ &\qquad \text{for} \quad \tau_1 \geqq 0\end{aligned} \tag{26}$$

$$\begin{aligned}\int_0^\infty\int_0^\infty h_2(\alpha,\beta)&\overline{x(t-\alpha)x(t-\beta)x(t-\tau_1)x(t-\tau_2)}\,d\alpha\,d\beta \\ &= \overline{x_\mathrm{d}(t)x(t-\tau_1)x(t-\tau_2)} - h_0\,\overline{x(t-\tau_1)x(t-\tau_2)} \\ &\quad - \int_0^\infty h_1(\alpha)\overline{x(t-\alpha)x(t-\tau_1)x(t-\tau_2)}\,d\alpha \\ &\qquad \text{for} \quad \tau_1, \tau_2 \geqq 0\end{aligned} \tag{27}$$

The equations are derived by using the calculus of variations and can be shown to correspond to a minimum.

These equations are a set of integral equations which are linear in the unknown kernels. This point cannot be overemphasized as the design of a nonlinear filter of degree n is reduced to the solution of n linear integral equations.

As a result of the condition τ_1, $\tau_2 \geqq 0$ and the form of the integral operators, the equations resemble multidimensional Wiener-Hopf equations. But, as was indicated previously, the integrals are not multidimensional convolution integrals and the problem cannot be solved by transform methods.

From the equations it is seen that a first-degree filter is specified by the autocorrelation of the input and the crosscorrelation between the signal x and the desired output x_d. A filter of degree n is specified by giving the first $2n$ autocorrelation functions, $\overline{x}$, $\overline{xx}$, $\overline{xxx}$, $\cdots$, and n crosscorrelations between the input and the desired output, $\overline{x_d}$, $\overline{x_d x}$, $\overline{x_d xx}$, $\cdots$ $\overline{x_d x \cdots x}$.

It is interesting to note the following property of the equations. Let us assume that, in Fig. 4, $h_0 = k_0$ and $h_1 = k_1$ are given and it is required to find the optimum second-order kernel h_2 which minimize the mean-square error between a desired output x_d and the output of the filter. By applying the variational technique to minimize the mean-square error there results

$$\int_0^\infty \int_0^\infty h_1(\alpha, \beta)\overline{x(t-\alpha)x(t-\beta)x(t-\tau_1)x(t-\tau_2)}\, d\alpha\, d\beta$$
$$= \overline{x_d(t)x(t-\tau_1)x(t-\tau_2)} - k_0\, \overline{x(t-\tau_1)x(t-\tau_2)}$$
$$- \int_0^\infty k_1(\alpha)\overline{x(t-\alpha)x(t-\tau_1)x(t-\tau_2)}\, d\alpha$$
$$\text{for} \quad \tau_1, \tau_2 \geqq 0$$

which is the same as Eq. (27) except that k_0 and k_1 replace h_0 and h_1. Therefore, the equation for the optimum h_j where the kernels h_i, $i = 1 \cdots n, j - 1$, are given is the same as the jth equation in the system of equations which determine the optimum filter of order j. This property will be used later for solving the set of equations for the optimum filter of degree n.

Let

$$\epsilon_{01}(t) = x_d(t) - k_0 - \int_0^\infty k_1(\alpha)x(t - \alpha)\, d\alpha$$

ϵ_{01} is the error which remains after filtering with k_0 and k_1. The equation for the optimum $h_2(\alpha, \beta)$ becomes

$$\int_0^\infty \int_0^\infty h_2(\alpha, \beta)\overline{x(t - \alpha)x(t - \beta)x(t - \tau_1)x(t - \tau_2)}\, d\alpha\, d\beta$$

$$= \overline{\epsilon_{01}(t)x(t - \tau_1)x(t - \tau_2)}$$

$$\text{for} \quad \tau_1, \tau_2 \geqq 0$$

which can be solved by the use of eigenfunctions and eigenvalues as outlined above.

In the integrals which appear on the left-hand side of Eqs. (25), (26), and (27) all kernels are symmetric in all variables as they are even-order autocorrelation functions. Under the proper assumptions (see previous section), all the kernels are bounded and can be made square integrable by attenuating the past in the same way as was done previously. This means that the filters $h_1(\alpha)$ and $h_2(\alpha, \beta)$ operate on $x(t - \alpha)e^{-k\alpha}$, $x(t - \beta)e^{-k\beta}$, $(k > 0)$ respectively, instead of operating on $x(t - \alpha)$ and $x(t - \beta)$. The following shorthand notation will be used:

$$\overline{xx} = 1(\alpha)1(t_1)\overline{x(t - \alpha)x(t - t_1)}\, e^{-k(\alpha+t_1)}$$

$$\overline{xxxx} = 1(\alpha)1(\beta)1(t_1)1(t_2)x(t - \alpha)x(t - \beta)x(t - t_1)$$
$$\cdot x(t - t_2)e^{-k(\alpha+\beta+t_1+t_2)}$$

$$\overline{xxx} = 1(\alpha)1(\beta)1(t_1)\overline{x(t - \alpha)x(t - \beta)x(t - t_1)}\, e^{-k(\alpha+\beta+t_1)}$$

$$\overline{x_d\, x} = 1(t_1)\overline{x_d(t)x(t - t_1)}\, e^{-kt_1}$$

$$\overline{x_d\, xx} = 1(t_1)1(t_2)\overline{x_d(t)x(t - t_1)x(t - t_2)}\, e^{-k(t_1+t_2)}$$

$$\int h_1\, \phi\, d\alpha = \int_0^\infty h_1(\alpha)\phi(\alpha)\, d\alpha$$

$$\iint h_2\, \phi\, d\alpha\, d\beta = \int_0^\infty \int_0^\infty h_2(\alpha, \beta)\phi(\alpha, \beta)\, d\alpha\, d\beta$$

Using the above notation Eqs. (25), (26), and (27) become:

$$h_0 = \overline{x_d} - \int h_1 \overline{x} \, d\alpha - \iint h_2 \overline{xx} \, d\alpha \, d\beta \tag{28}$$

$$\int h_1 \overline{xx} \, d\alpha = \overline{x_d x} - h_0 \overline{x} - \iint h_2 \overline{xxx} \, d\alpha \, d\beta \tag{29}$$

$$\iint h_2 \overline{xxxx} \, d\alpha \, d\beta = \overline{x_d \, xx} - h_0 \overline{xx} - \int h_1 \overline{xxx} \, d\alpha \tag{30}$$

Before actually solving the set of equations let us review a few properties. The kernels $\overline{xx}$, $\overline{xxxx}$ of the integrals on the left-hand side of the equations are symmetric, bounded and square integrable in each and all variables. Therefore each possesses a set of eigenvalues and corresponding eigenfunctions (Courant and Hilbert, 1953; Smithies, 1958). All eigenvalues are positive since the kernels are positive definite. The functions appearing on the right-hand side of the set of equations can be specified in the mean by using only the set of eigenfunctions of the kernels which appear on the left-hand side of the corresponding equation. (The proof is the same as in Appendix A.) Any of the above equations, when the right-hand side function is given, can be solved by using the method of eigenfunctions and eigenvalues which was developed above. In the following section the above properties are used for developing an iteration method for solving the set of integral equations.

ITERATION METHOD FOR SOLUTION OF THE INTEGRAL EQUATIONS

In this section an iteration process for solving the filter's equations is developed. The iteration is done in a way that insures a physical meaning to the result after each iteration step. The iteration can be terminated at any step yielding an optimum filter in some sense. The result of the iteration process converges to the optimum filter of degree n.

As before, a filter of the second degree is considered. The equations to be solved are Eqs. (28), (29), and (30).

First stage of the iteration process:

1. We start the iteration process by asking: What is the optimum filter of zero degree, h_0 ? As was already mentioned, Eq. (28) specifies the optimum h_0 when h_1 and h_2 are given. The above question is answered by solving Eq. (28) when $h_1 = 0$, $h_2 = 0$. The equation becomes

$$h_0^1 = \overline{x_d} \tag{31}$$

where the superscript 1 denotes that h_0^1 is the result of the first stage of the iteration process. Let us denote the error at the output of h_0^1 by ϵ_0^1.

2. The second step answers the question: What is the optimum filter of the first degree which is to be connected in parallel with h_0^1? This filter is specified by Eq. (29) in which h_0 is taken equal to h_0^1 and $h_2 = 0$. This first-order filter is denoted by h_1^1 and the equation becomes:

$$\int h_1^1 \overline{xx}\, d\alpha = \overline{x_d x} - h_0^1 \overline{x} \tag{32}$$

The solution of Eq. (32) yields the linear filter h_1^1 which is the Wiener optimum filter when $\overline{x_d} = 0$. Let us denote the error at the output of the filter which consists of h_0^1 and h_1^1 by ϵ_1^1.

3. The third step of the first stage is to find the optimum second-order filter h_2^1 when h_0^1 and h_1^1 are given. The filter is given by Eq. (30) when $h_0 = h_0^1$; $h_1 = h_1^1$

$$\iint h_2^1 \overline{xxxx}\, d\alpha\, d\beta = \overline{x_d xx} - h_0^1 \overline{xx} - \int h_1^1 \overline{xxx}\, d\alpha \tag{33}$$

As before we denote the error at the output of the filter which consists of h_0^1, h_1^1 by ϵ_2^1.

Second stage of the iteration process: The results of the first stage are used as a starting point for improvement.

In the first step one finds the best zero order filter, h_0^2, when h_1^1 and h_2^1 are given. The equation is

$$h_0^2 = \overline{x_d} - \int h_1^1 \overline{x}\, d\alpha - \iint h_2^1 \overline{xx}\, d\alpha\, d\beta \tag{34}$$

The error at the output of $\langle h_0^2, h_1^1, h_2^1 \rangle$ is denoted by ϵ_0^2.

The second step yields the best linear filter when h_0^2 and h_2^1 are present. The equation is

$$\int h_1 \overline{xx}\, d\alpha = \overline{x_d x} - h_0^2 \overline{x} - \iint h_2^1 \overline{xxx}\, d\alpha\, d\beta \tag{35}$$

The error at the output is denoted by ϵ_1^2. Similarly h_2^2 and ϵ_2^2 result from

$$\iint h_2^2 \overline{xxxx}\, d\alpha\, d\beta = \overline{x_d xx} - h_0^2 \overline{xx} - \int h_1^2 \overline{xxx}\, d\alpha \tag{36}$$

Continuing in this form the equations at the ith stage are:

$$h_0^i = \overline{x_d} - \int h_1^{i-1}\overline{x}\, d\alpha - \iint h_2^{i-1}\overline{xx}\, d\alpha\, d\beta \tag{37}$$

$$\int h_1^i\overline{xx}\, d\alpha = \overline{x_d x} - h_0^i\overline{x} \iint h_2^{i-1}\overline{xxx}\, d\alpha\, d\beta \tag{38}$$

$$\iint h_2^i\overline{xxxx}\, d\alpha\, d\beta = \overline{x_d xx} - \int h_0^i\, \overline{xx} - h_1^i\overline{xxx}\, d\alpha \tag{39}$$

The iteration results in a series of filters and values for the corresponding mean-square errors:

$$h_0^1, h_1^1, h_2^1, h_0^2, h_1^2, h_2^2, h_0^3, h_1^3, h_2^3, h_0^4 \cdots$$

$$\overline{(\epsilon_0^1)^2}, \overline{(\epsilon_1^1)^2}, \overline{(\epsilon_2^1)^2}, \overline{(\epsilon_0^2)^2}, \overline{(\epsilon_1^2)^2}, \overline{(\epsilon_2^2)^2}, \overline{(\epsilon_0^3)^2}, \overline{(\epsilon_1^3)^2}, \overline{(\epsilon_2^3)^2}, \overline{(\epsilon_0^4)^2} \cdots$$

Every three adjoining terms in the h_1 series have the property that the last one is the optimum filter when the two previous ones are present. The $\overline{\epsilon^2}$ corresponding to the third filter in the three-term subsequence is the mean-square error of the system represented by the three terms. Convergence of the iteration process is discussed in Appendix C. It can also be shown that the solution is unique up to a filter $\langle k_0, k_1, k_2 \rangle$ which does not effect the mean square error. The proof of uniqueness is given in Appendix B.

Another way of performing the iteration process uses the linearity of the integral equations in the following way. Let

$$\begin{aligned} \Delta h_0^i &= h_0^i - h_0^{i-1} \\ \Delta h_1^i &= h_1^i - h_1^{i-1} \\ \Delta h_0^i &= h_2^i - h_2^{i-1} \end{aligned} \tag{40}$$

with the agreement that

$$h_0^0 = h_1^0 = h_2^0 \triangleq 0$$

By substracting Eqs. (37), (38), and (39) for $i = j - 1$ from the corresponding equations for $i = j$ and using the notations (40) we get for $j > 1$

$$\Delta h_0^j = -\int \Delta h_1^{j-1}\, \overline{x}\, d\alpha - \iint \Delta h_2^{j-1}\overline{xx}\, d\alpha\, d\beta \tag{41}$$

$$\int \Delta h_1^j\overline{xx}\, d\alpha = -\Delta h_0^j\overline{x} - \iint \Delta h_2^{j-1}\overline{xxx}\, d\alpha\, d\beta \tag{42}$$

$$\iint \Delta h_2^j\overline{xxxx}\, d\alpha\, d\beta = -\Delta h_0^j\overline{xx} - \int \Delta h_1^j\overline{xxx}\, d\alpha \tag{43}$$

Using these equations, the stages in the iteration process are as follows. The first stage is the same as in the former approach. From Eqs. (31), (32), and (33), h_0^1, h_1^1, and h_2^1 are found. In the second stage, the following question is answered: What filter, Δh_0^1, is to be added to h_0 in order that $h_0 + \Delta h_0^1$ be the optimum zero-order filter when h_1^1 and h_2^1 are present? This question is answered by solving Eq. (41) for $j = 1$ and getting Δh_0^1. Similar questions lead to Eqs. (42) and (43) and to Δh_1^1 and Δh_2^1. The following stages are similar and use the Δh^{i-1} to find Δh^i. This iteration leads to the series

$$h_0^1(\Delta h_0^0);\ h_1^1(\Delta h_1^0);\ h_2^1(\Delta h_2^0);\ \Delta h_1^1;\ \Delta h_2^1;\ \Delta h_0^2;\ \Delta h_1^2;\ \Delta h_2^2\ \cdots$$

where the optimum filter is given by

$$h_0 = \sum_{i=0}^{\infty} \Delta h_0^i$$

$$h_1 = \sum_{i=0}^{\infty} \Delta h_1^i$$

$$h_2 = \sum_{i=0}^{\infty} \Delta h_2^i \tag{44}$$

One can look at this method as a procedure which, at each step, adds a correction of a suitable degree to an existing filter. The sum of the successive corrections is the optimum filter.

Strictly speaking, the two methods are equivalent. Technically, both have the advantage that the same eigenfunctions of the kernels $\overline{xx}$, $\overline{xxxx}$ appear in all stages of the iteration. Therefore they are computed in the beginning once and for all. Practically, the second method might be more convenient as the rate of decrease of Δh_1^k might indicate how good an approximation has been achieved.

It has to be emphasized that in both methods one does not have to start with h_0 to calculate h_1 and continue by calculating h_2, etc. At each stage any kernel can be calculated first and the fact that in a previous stage the order was different is irrelevant.

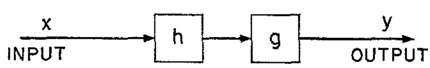

FIG. 5. A cascade compensation network

EXPRESSIONS FOR THE REMAINING ERROR

The mean-square error when the optimum $\langle h_0, h_1, h_2 \rangle$ is used is

$$\overline{\epsilon^2} = E\left\{\left(x_d - h_0 - \int_0^\infty h_1(\alpha)x(t-\alpha)\,d\alpha - \int_0^\infty \int_0^\infty h_1(\alpha,\beta)x(t-\alpha)x(t-\beta)\,d\alpha\,d\beta\right)^2\right\}$$

$$= \overline{x_d^2} + \overline{h_0^2} + \int h_1 \int h_1 \overline{xx}\,d\alpha\,d\beta + \iint h_2 \iint h_2 \overline{xxxx}\,d\alpha\,d\beta\,d\alpha\,d\beta - 2h_0\overline{x_d} - 2h_0 \int h_1 \overline{x}\,d\alpha - 2h_0 \iint h_2 \overline{xx}\,d\alpha\,d\beta - 2\int h_1 \overline{xx_d}\,d\alpha - 2\int h_1 \iint h_2 \overline{xxx}\,d\alpha\,d\beta\,d\alpha - 2\iint h_2 \overline{xxx_d}\,d\alpha\,d\beta \tag{45}$$

By using Eqs. (28), (29), and (30) we get

$$\overline{\epsilon^2} = \overline{x_d^2} - h_0\overline{x_d} - \int h_1 \overline{x_d x}\,d\alpha - \iint h_2 \overline{x_d xx}\,d\alpha\,d\beta \tag{46}$$

The expression is similar in form to the expression for the error in the cascade configuration which was discussed previously and to the expression of the error when the Wiener linear filter is used.

APPLICATION TO CONTROL

While the application of the above technique to prediction or filtering of signal from noise or combined prediction and filtering is straightforward, the design of nonlinear compensation networks for control systems requires some additional discussion.

The analytical design technique (Newton *et al.*, 1957) deals mainly with the control problem illustrated in Fig. 5 where g is a fixed network and h is a compensation network. It is required to find a linear compensation network h which optimizes the mean-square difference between the output and a desired output. The problem is stated in several forms all of which lead to more or less similar equations.

For a deterministic input x the desired output x_d is also deterministic and one has to find the best linear compensation to minimize

$$\bar{\bar{\epsilon}} = \int_0^\infty (x_d(t) - x(t))^2 \, dt$$

where $\bar{\bar{\epsilon}}$ is called the integral-square error. To insure convergence of the integral and to emphasize transient error over steady-state errors, a weighing function is often used and $\bar{\bar{\epsilon}}$ becomes

$$\bar{\bar{\epsilon}} = \int_0^\infty W(t)(x_d(t) - x(t))^2 \, dt$$

When the input is a stationary egodic random process, x_d is a related random process and the criterion will involve the expectation of the square of the difference $E(x_d(t) - y(t))^2$.

The fixed element g was constrained to be linear. In many problems a trivial result is obtained if g is invertable, that is, g^{-1} is a physically realizable network. The question has nontrivial mean when g is not a minimum phase network (Newton *et al.*, 1957) or when various constraints are used. Typical constraints are minimization of bandwidth and a constraint on the amplitude of the signal.

In the following sections the same compensation problem is treated. But instead of limiting the compensation network to be linear it is limited to be a nonlinear filter of order n. The fixed-element network has to remain linear. In fact, when the fixed element is not linear the equations can become so complicated that it seems to be impractical to solve them by the iteration method.

Free Configuration Compensation with a Deterministic Input

In Fig. 5 let us assume x to be a deterministic signal. Let us assume that $x(t)$ is an L^2 function on $(-\infty, 0)$. If $x(t)$ is not L^2, but is bounded, a convergence factor can be used. It is required to find $\langle h_0, h_1, h_2 \rangle$ which minimizes the expression

$$\bar{\bar{\epsilon}} = \int_0^\infty (x_d(t) - y(t))^2 \, dt$$

If we use the notation

$$\overline{\overline{f(x)}} = \int_0^\infty f(x) \, dx$$

and let

$$z_0 = \overline{\overline{g}}$$
$$z_1(t) = \overline{\overline{g(\alpha)x(t-\alpha)}}$$
$$z_2(t_1, t_2) = \overline{\overline{g(\alpha)x(t_1-\alpha)x(t_2-\alpha)}}$$

then the equations for $\langle h_0, h_1, h_2\rangle$ are identical to Eqs. (28), (29), and (30) except that expectations (which are denoted by a single bar) are replaced by integrals denoted by a double bar. The integrals have the same properties as the expectations if exponential weighting factors are used to make the kernels L^2. The presence of the fixed element has changed nothing in principle and the method of solution of the problem is the same as outlined previously. A similar result holds when x and x_d are sample functions of a stationary ergodic random process.

Fixed Element with Amplitude Constraints

Most systems and components used in practice cannot handle input amplitudes or input power larger than a certain amount. When designing compensation networks for such elements this characteristic cannot be overlooked and suitable constraints have to be placed on the input. However, with the mathematical technique which we are using, it is very difficult, or may be impossible, to include amplitude constraints directly. We can, however, lower the probability that the amplitude will exceed a certain value by constraining the mean square of this value (Newton *et al.*, 1957). The technique is best illustrated by considering Fig. 6 and the following example.

In Fig. 6, x is a random input which is a sample function from a stationary ergodic process; g is the fixed element; x_d is a random process related to x. It is required to design a cascade compensation of second degree, $\langle h_0, h_1, h_2\rangle$, to optimize $\overline{(x_d - y)^2}$. The optimization has to be done while the mean square of a quantity q, which is linearly related to the input of g, does not exceed a certain value, σ_0. The linear relation mentioned above appears in Fig. 6 as the linear filter k. When one wants to limit the input power, $k = 1$. When the output power is to be limited,

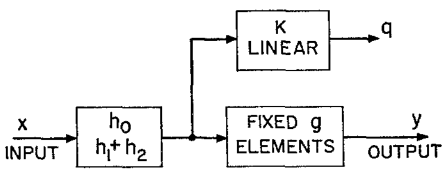

FIG. 6. Compensation with amplitude constraint

$k = g$. When the amplitude has to be limited, a suitable choice of σ_0 will maintain the probability of entering saturation below a certain value.

The problem to be solved is: Find $\langle h_0, h_1, h_2 \rangle$ which minimizes $\overline{(x_d - y)^2}$ subject to the constraint $\overline{q^2} < \sigma_0$. We use the Lagrange multiplier, ρ, and minimize

$$\overline{(x_d - y)^2} + \rho \overline{q^2}$$

This results in optimum filters $\langle h_{0\rho}, h_{1\rho}, h_{2\rho} \rangle$. For every ρ a different nonlinear compensation will result. Let us denote each of these filters by $H(\rho)$. Then a value for $\overline{q^2} = \overline{q^2}(\rho)$ will result. One chooses a ρ for which $\overline{q^2}(\rho) < \sigma_0$ and minimizes $\overline{\epsilon^2}$. The last step is tedious and usually one solves the equations for a few values of ρ and chooses the one which fulfills the requirements.

Let

$$z_1(t) = \int_0^\infty g(\alpha)x(t - \alpha)\, d\alpha$$

$$Z_1(t) = \int_0^\infty q(\alpha)x(t - \alpha)\, d\alpha$$

$$z_2(t_1, t_2) = \int_0^\infty g(\alpha)x(t - \alpha - t_1)x(t - \alpha - t_2)\, d\alpha$$

$$Z_2(t_1, t_2) = \int_0^\infty k(\alpha)x(t - \alpha - t_1)x(t - \alpha - t_2)\, d\alpha$$

$$z_0 = \int_0^\infty g(\alpha)\, d\alpha$$

$$Z_0 = \int_0^\infty k(\alpha)\, d\alpha$$

After applying the standard variational technique the equations for $\langle h_0, h_1, h_2 \rangle$ become (in shorthand notation):

$$h_0(1 + \rho) = \overline{x_d z_0} - \int_0^\infty h_1\overline{(z_1 + \rho Z_1)}\, d\alpha - \int_0^\infty \int_0^\infty h_2\overline{(z_2 + \rho Z_2)}\, d\alpha\, d\beta$$

$$\int_0^\infty h_1\overline{(z_1 z_1} + \rho \overline{Z_1 Z_1})\, d\alpha = \overline{x_d z_1} - h_0\overline{(z_0 z_1} + \rho \overline{Z_0 Z_1})$$
$$- \int_0^\infty \int_0^\infty h_2\overline{(z_2 z_1} + \rho \overline{Z_2 Z_1})\, d\alpha\, d\beta \qquad \text{for } t_1 > 0$$

$$\int_0^\infty \int_0^\infty h_2\overline{(z_2 z_2} + \rho \overline{Z_2 Z_2})\, d\alpha\, d\beta = \overline{x_d z_2} - h_0\overline{(z_0 z_2} + \rho \overline{Z_0 Z_2})$$
$$- \int_0^\infty h_1 \left(\overline{z_1 z_2} + \rho \overline{Z_1 Z_2}\right) d\alpha \qquad \text{for } t_1, t_2 > 0$$

All the kernels appearing on the left are symmetric and can be made (if they are not) L^2, and the equations can be solved in the usual way. In order to find ρ, the results are treated as explained in the beginning of the section.

Nonlinear Fixed Element

In this section we shall determine the equations which define an optimum compensation network for a nonlinear fixed element. The example treated is instructive since it illustrates the limitation of the method of this work.

In Fig. 5 let the fixed element be a second-order filter $\langle 0, g_1, g_2 \rangle$. It is required to design an optimum compensation which is a linear filter, h_1. As usual x and x_d are related stationarily ergodic random processes.

Let

$$z_1^1 = \int_0^\infty g_1(\gamma) x(t - \alpha - \gamma)\, d\beta$$

$$z_2^1 = \int_0^\infty \int_0^\infty g_2(\sigma) x(t - \alpha - \gamma) x(t - \beta - \sigma)\, d\gamma\, d\sigma$$

$$z_2^2 = \int_0^\infty \int_0^\infty g_2(\gamma, \sigma) x(t - \alpha - \gamma) x(t - \beta - \gamma) x(t - \epsilon - \sigma)\, d\gamma\, d\sigma$$

Then by applying the standard variational technique the integral equation for h_1 is

$$\begin{aligned}\overline{x_d\, z_1^1} + \int_0^\infty h_1(2\overline{z_2^1 x_d} + \overline{z_1^1 z_1^1})\, d\beta + 3 \int_0^\infty \int_0^\infty h_1\, h_1\, \overline{z_1^1 z_2^1}\, d\beta\, d\gamma \\ + 2 \int_0^\infty \int_0^\infty \int_0^\infty h_1\, h_1\, h_1 [\overline{z_2^1 z_2^1} + 2\overline{z_2^1 z_2^2}]\, d\beta\, d\gamma\, d\epsilon = 0 \qquad \text{for } t_1 > 0\end{aligned} \tag{48}$$

The above integral equation is no longer linear. In fact in this specific example the unknown kernel appears to the third "power." Equations which involve nonlinear compensation are even more complicated. It might be that equations of this degree of complexity could be solved. Probably, the solution would be much more complicated than the solution of the linear integral equation in the same way as the solution of an algebraic equation of higher order is more difficult than that of a linear equation.

This example illustrates the following point. Although it is not claimed that equations like (48) cannot be solved, it is clear that they are too complicated to be solved by our method. The eigenfunction method and the iteration process requires the unknown kernel h_i to appear in the ith equation to the first power, as the proof of convergence is based on this fact. As this is not the case here, our method does not apply.

One thus can conclude that the fixed element has to be linear otherwise the equation becomes too complicated.

SUMMARY AND CONCLUSIONS

This paper describes a systematic analytic approach to nonlinear filtering which can be used for any stationary random process provided certain averages exist. Instead of looking for the optimum possible filter for a given random process the method uses a filter of a given form as an approximation to the optimum possible continuous filter. The filter used is a filter of degree n—a filter which is described by the first n Volterra kernels. This approximation has the following properties: The accuracy of the approximation increases with the number of kernels involved and its performance approaches arbitrarily close to the performance of the optimum possible continuous filter. The advantage of the particular form of approximation is that it starts with the simple elements first and adds more and more complicated elements as the accuracy requirements are increased.

The statistical data which specify the filter are the higher order auto- and crosscorrelation functions. A filter of degree n is given by the first $2n$ autocorrelation and the first n crosscorrelation functions. In some cases conclusions as to the form of the filter can be derived from the form of the correlation functions.

By choosing a specific canonical form to express the functional which corresponds to the nonlinear filter one actually gives up the convenience of using a canonical expansion in terms of functionals which are orthogonal to each other. This results in equations which are much more complicated than those which would have corresponded to an orthogonal expansion. The main advantage of this choice is that the method is not dependent upon the specific random process in question.

The filter of degree n is specified by a set of n simultaneous integral equations which are linear in the unknown kernels. This equation is solved by first solving the filtering problem when a single Volterra kernel is involved and then an iteration method is used to extend the solution to a filter of degree n.

Part II of this paper will discuss the applicability of this method for analytic solution of given problems and for experimental calculation of nonlinear filters. The main conclusions that will be demonstrated are that problems can be solved analytically although the complication increases very rapidly with the degree of the kernel. As the gaussian process is the only one for which there exists a complete statistical description in terms of autocorrelation analytical treatment is limited, for the present, to gaussian-derived processes. The experimental characterization of the pupil of the human eye shows that the method can be applied to real problems. The application is fairly complicated and each problem has to be checked to estimate the complexity of the solution.

APPENDIX A. SOLUTION OF THE INTEGRAL EQUATION

$$f(t) = \int_\Omega K(t, s)h(s)\, ds \tag{A.1}$$

where $f(t)$ is a bounded, L^2 function and $K(s, t)$ is symmetric and L^2 in each and both s and t, that is

$$\int_\Omega K^2(s, t)\, dt < \infty, \text{ all } s \text{ in } \Omega; \qquad \int_\Omega\int_\Omega K^2(s, t)\, ds\, dt < \infty$$

Having the above properties, $K(s, t)$ possesses a set of eigenfunctions ψ_i and eigenvalues λ_i (Courant and Hilbert, 1953; Smithies, 1958). The eigenfunctions are a normalized orthogonal set of functions which with the corresponding eigenvalues obey the integral equation

$$\lambda_i \int_\Omega K(s, t)\psi_i(s)\, ds = \psi_i(t). \tag{A.2}$$

In the domain Ω, $K(s, t)$ possesses an orthogonal expansion in terms of its eigenfunctions and eigenvalues

$$K(s, t) \overset{\circ}{=} \sum_{i=1}^{\infty} \frac{\psi_i(s)\psi_i(t)}{\lambda_i} \tag{A.3}$$

In this series the λ_i are arranged according to increasing magnitude of their absolute value $|\lambda_1| \leqq |\lambda_2| \leqq |\lambda_3|, \cdots$. The sign $\overset{\circ}{=}$ means that the right-hand-side series approaches $K(s, t)$ in the mean as n tends to infinity. The limit-in-the-mean becomes an equality if the kernel is degenerate, that is, if it possesses a finite number of eigenfunctions and eigenvalues, or when the kernel is continuous and definite (Mercer theorem). Let $\{\phi_i\}$ be a complete set of normalized orthogonal functions on Ω such that the members of the set of eigenfunctions of $K(s, t)$, $\{\psi_i\}$, are members of $\{\phi_i\}$. As $f(t)$ is L^2 on Ω it can be expanded in terms of $\phi_i(t)$

$$f(t) \overset{\circ}{=} \sum_{i=1}^{\infty} (f, \phi_i)\phi_i(t) \tag{A.4}$$

where

$$(f, \phi_i) = \int_\Omega f(t)\phi_i(t)\, dt \tag{A.5}$$

We will assume that $f(t)$ can be completely specified by those members of $\{\phi_i\}$ which are eigenfunctions of $K(s, t)$, $\{\psi_i\}$. This means that

$$\int_\Omega f(t)\phi_i(t)\, dt = 0$$

when ϕ_i is not a member of $\{\psi_i\}$ and Eq. (A.4) can be written as

$$f(t) \overset{\circ}{=} \sum_{i=1}^{\infty} (f, \psi_i)\psi_i(t) \tag{A.6}$$

The justification of this assumption will be shown later.

Let us define a function $h_n(s)$ by

$$h_n(s) = \sum_{i=1}^{n} \lambda_i(f, \psi_i)\psi_i(s) \tag{A.7}$$

$h_n(s)$ is a solution of Eq. (A.1) in the following sense;

$$f(t) = \underset{n\to\infty}{\text{l.i.m.}} \int_\Omega K(t, s)h_n(s)\, ds; \tag{A.8}$$

The proof of this follows.

$$\begin{aligned}
\int_\Omega (f(t) - \int_\Omega K(t, s)h_n(s)\, ds)^2\, dt \\
&= \int_\Omega f^2(t)\, dt - 2\int_\Omega\int_\Omega K(t, s)f(t)h_n(s)\, dt\, ds \\
&\qquad + \int_\Omega \left[\int_\Omega K(s, t)h_n(s)\, ds\right]^2 dt \\
&= \int_\Omega f^2(t)\, dt - 2\int_\Omega\int_\Omega K(s, t)f(t) \sum_{i=1}^{\infty} \lambda_i(f, \psi_i)\psi_i(s)\, ds\, dt \\
&\qquad + \int_\Omega \left[\int_\Omega K(s, t) \sum_{i=1}^{n} \lambda_i(f, \psi_i)\psi_i(s)\, ds\right]^2 dt \\
&= \int_\Omega f^2(t)\, dt - \sum_1^n (f, \psi_i)^2
\end{aligned} \tag{A.9}$$

From Eq. (A.6) and Bessel's equality

$$\int_{\Omega} f^2(t)\, dt = \sum_{i=1}^{\infty} (f, \psi_i)^2 \tag{A.10}$$

Using Eq. (A.10) and Eq. (A.9) we get

$$f(t) = \underset{n\to\infty}{\text{l.i.m.}} \int_{\Omega} K(t, s) h_n(s)\, ds \tag{A.11}$$

Before claiming that the equation is solved, the assumption that in our problem $f(t)$ can be specified in the mean by the eigenfunctions of $K(s, t)$ has to be investigated.

For any filter h the mean-square error is

$$\overline{\epsilon_2^2} = \overline{\epsilon_1^2} - 2 \int_{\Omega} h(t) f(t)\, dt + \int_{\Omega}\int_{\Omega} K(s, t) h(t) h(s)\, dt\, ds \tag{A.12}$$

As a quadratic form $\overline{\epsilon_2^2} \geqq 0$ for all h. Let $\bar{\phi}(t)$ be a normalized function orthogonal to all eigenfunctions of $K(s, t)$. Let us assume that

$$\int_{\Omega} f(t)\bar{\phi}(t)\, dt \neq 0 \tag{A.13}$$

Let us choose h to be $h(t) = a \cdot \bar{\phi}(t)$, where a is a constant. The mean-square error will be

$$\begin{aligned} \overline{\epsilon_2^2} &= \overline{\epsilon_1^2} - 2a(f, \bar{\phi}) + a^2 \int_{\Omega}\int_{\Omega} K(s, t)\bar{\phi}(t)\bar{\phi}(s(\, ds\, dt \\ &= \overline{\epsilon_1^2} - 2a(f, \bar{\phi}) \end{aligned} \tag{A.14}$$

Now, by choosing a suitable a, the above expression can be made negative, or $\overline{\epsilon_2^2} < 0$. From this it follows that $\bar{\phi}(t)$ does not exist, or, $f(t)$ can be specified in the mean by the eigenfunctions of $K(s, t)$.

APPENDIX B. CONVERGENCE OF THE ITERATION PROCESS AND UNIQUENESS OF THE SOLUTION

It will be proved that the iteration process converges to a solution of the equations which is unique except for an operator which does not effect the error.

Let us consider the series of filters which result from the iteration process and the corresponding values of the mean-square error.

$$h_0^{1}, h_1^{1}, h_2^{1}, h_0^{2}, h_1^{2}, h_2^{2}, h_0^{3}, h_1^{3}, h_2^{3}, \cdots, h_0^{j}, h_1^{j}, h_2^{j} \tag{B.1}$$

$$\epsilon_1^{0}, \epsilon_1^{1}, \epsilon_1^{2}, \epsilon_2^{1}\epsilon_2^{2}\epsilon_2^{3}; \epsilon_3^{0}\epsilon_3^{1}\epsilon_3^{2}; \cdots, \epsilon_j^{0}\epsilon_j^{1}\epsilon_j^{2}; \cdots \tag{B.2}$$

As a result of the method in which the series (B.1) was constructed, the filter $\langle h_0^{j+1}, h_1^{j+1}, h_2^{j+1}\rangle$ is better than or equal in performance to $\langle h_0^j, h_1^j, h_2^j\rangle$. Any filter which is constructed of three adjoining expressions of Eq. (B.1) is better than or equal in performance to any filter which is constructed from h_i's which appear in Eq. (B.1) in an earlier position. This means that

$$\epsilon_j^k \geqq \epsilon_i^l \qquad \text{if } i \geqq j \quad \text{and} \quad l > k$$

or the series (B.2) is monotonic decreasing. As the mean-square error is positive, $\epsilon_j^k \geqq 0$ (all k and j) and the series is bounded from below. As the series is monotonic and bounded the limit

$$\lim_{i\to\infty} \epsilon_i^{0,1,2}$$

exists. Let us denote the upper lower bound of the series by ϵ

$$\epsilon = \lim_{i\to\infty} \epsilon_i^{0,1,2}$$

Let $\langle h_0, h_1, h_2\rangle$ be the filter which corresponds to the error ϵ. $\langle h_0, h_1, h_2\rangle$ will be

$$h_0 = \lim_{j\to\infty} h_0^j; \qquad h_1 = \lim_{j\to\infty} h_1^j; \qquad h_2 = \lim_{j\to\infty} h_2^j \qquad \text{(B.3)}$$

We claim that $\langle h_0, h_1, h_2\rangle$ are solutions of the integral equations.

In order to prove this, let us first consider another property of the series (B.1) and (B.2).

By using notation (40) and the expression for the error (46), we get the following relations between errors which correspond to two successive iteration steps.

$$\epsilon_k^0 - \epsilon_{k-1}^2 = -(\Delta h_0^k)^2 \leq 0$$

$$\epsilon_k^1 - \epsilon_k^0 = -\overline{(\textstyle\int \Delta h_1^k x)}^2 \leq 0$$

$$\epsilon_k^2 - \epsilon_k^1 = -\overline{(\textstyle\int \Delta h_2^k xx)}^2 \leq 0$$

The expression for the filter is of the form

$$\Delta h = \sum_1^\infty a_i \psi_i$$

where ψ_i are the eigenfunctions of the corresponding autocorrelation function. Thus, the improvement in the mean square error at each iteration step is of the form

$$\Delta\epsilon = \sum \frac{a_i^2}{\lambda_i}$$

where the λ_i are the corresponding eigenvalues. As the eigenvalues are positive, $\Delta\epsilon = 0$ only if all a_i are zero.

Thus, an iteration stage cannot cause changes in the filter which do not improve the performance. The filter is either improved or remains unchanged.

Now, assume that $\langle h_0, h_1, h_2 \rangle$ is not a solution of the equations. Then h_1 and h_2 can be assumed as given and by using the iteration process, Eq. (28), one can find k_0, the optimum filter of zero degree, when h_1 and h_2 are given. Now, if k_0 is not equal to h_0, then by this process some improvement is achieved and the error corresponding to $\langle k_0, h_1, h_2 \rangle$ is smaller than ϵ. However, ϵ is the upper lower bound of the iteration series (B.2). This leads us to a contradiction and implies that $k_0 = h_0$ and $\langle h_0, h_1, h_2 \rangle$ are solutions of the first equation. We can treat Eqs. (38) and (39) similarly and find that $\langle h_0, h_1, h_2 \rangle$ is the solution of the equations. Or

$$\lim_{i\to\infty} h_0^i = \lim_{i\to\infty} \left\{ \overline{x_d} - \int h_1^{i-1}\, \bar{x}\, d\alpha - \iint h_2^{i-1}\, \overline{xx}\, d\alpha\, d\beta \right\} \tag{B.4}$$

$$\lim_{i\to\infty} \int h_1^i \overline{xx}\, d\alpha = \lim_{i\to\infty} \left\{ \overline{x_d x} - h_0^i \bar{x} - \iint h_2^{i-1} \overline{xxx}\, d\alpha\, d\beta \right\} \tag{B.5}$$

$$\lim_{i\to\infty} \iint h_2^i \overline{xxxx}\, d\alpha\, d\beta = \lim_{i\to\infty} \left\{ \overline{x_d xx} - h_0^i \overline{xx} - \int h_1^i \overline{xxx}\, d\alpha \right\} \tag{B.6}$$

The result of the iteration process is unique in the sense that any other solution of the equation does not improve the error.

Let $\langle h_0, h_1, h_2 \rangle$ be the result of the iteration process and $\langle k_0, k_1, k_2 \rangle$ another filter which fulfills Eqs. (28), (29), and (30). We shall prove that the error corresponding to $\langle k_0, k_1, k_2 \rangle$ is equal to that which corresponds to $\langle h_0, h_1, h_2 \rangle$.

Let

$$\begin{aligned} \Delta h_0 &= k_0 - h_0 \\ \Delta h_1 &= k_1 - h_1 \\ \Delta h_2 &= k_2 - h_2 \end{aligned} \tag{B.7}$$

The mean square error when $\langle k_0, k_1, k_2 \rangle$ is used is

$$\overline{\epsilon_k^2} = \overline{x_d^2} - k_0 \overline{x_d} - \int k_1 \overline{x x_d}\, d\alpha - \iint k_2 \overline{x x x_d}\, d\alpha\, d\beta \tag{B.8}$$

From Eqs. (28), (29), and (30) for $\langle k_0, k_1, k_2 \rangle$

$$\overline{x_d} = k_0 + \int k_1 \overline{x}\, d\alpha + \iint k_2 \overline{xx}\, d\alpha\, d\beta$$

$$\overline{x x_d} = k_0 \overline{x} + \int k_1 \overline{xx}\, d\alpha + \iint k_2 \overline{xxx}\, d\alpha\, d\beta$$

$$\overline{x x x_d} = k_0 \overline{xx} + \int k_1 \overline{xxx}\, d\alpha + \iint k_2 \overline{xxxx}\, d\alpha\, d\beta$$

Putting these values in Eq. (46) we get

$$\overline{\epsilon^2} = \overline{x_d^2} - k_0 k_0 - \int k_1 \int k_1 \overline{xx}\, d\alpha\, d\beta \iint k_2 \iint k_2 \overline{xxxx}\, d\alpha\, d\beta\, d\gamma\, d\delta$$

$$- k_0 \int k_1 \overline{x}\, d\alpha - 2 \int k_1 \iint k_2 \overline{xxx}\, d\alpha\, d\beta\, d\gamma - 2k_0 \iint k_2 \overline{xx}\, d\alpha\, d\beta$$

By using notations from Eq. (B.7) we get (dropping the dx notation)

$$\overline{\epsilon^2} = \overline{x_d^2} - h_0 h_0 + 2\Delta h_0 h_0 + \Delta h_0^2 - \int h_1 \int h_1 \overline{xx}$$

$$+ 2 \int \Delta h_1 \int h_1 \overline{xx} + \int \Delta h_1 \int h_1 \overline{xx} - \iint h_2 \iint h_2 \overline{xxxx}$$

$$+ 2 \iint \Delta h_2 \iint h_2 \overline{xxxx} - 2 \int (h_0 h_1 + \Delta h_0 h_1 + h_0 \Delta h_1$$

$$+ \Delta h_1 \Delta h_0)\overline{x}\, d\alpha - 2 \iiint h_1 h_2 + \Delta h_1 h_2 + h_1 \Delta h_2$$

$$+ \Delta h_1 \Delta h_2)\overline{xxx}\, d\alpha\, d\beta\, d\gamma - 2 \iint (h_0 h_2 + \Delta h_0 h_2 + h_0 \Delta h_2$$

$$+ \Delta h_0 \Delta h_2)\overline{xx}\, d\alpha\, d\beta$$

The Δh_0, Δh_1, Δh_2 are related by the following equations which are derived by subtracting Eqs. (28), (29), and (30) for $\langle h_0, h_1, h_2 \rangle$ from the same equations for $\langle k_0, k_1, k_2 \rangle$

$$0 = \Delta h_0 + \int \Delta h_1\, x\, d\alpha + \iint \Delta h_2\, xx\, d\alpha\, d\beta$$

$$0 = \Delta h_0 + \int \Delta h_1\, xx\, d\alpha + \iint \Delta h_2\, xxx\, d\alpha\, d\beta$$

$$0 = \Delta h_0\, xx + \int \Delta h_1\, xxx\, d\alpha + \iint \Delta h_2\, xxxx\, d\alpha\, d\beta$$

Using these equations and once again Eqs. (28), (29), and (30) for $\langle h_0 h_1 h_2 \rangle$ we get

$$\overline{\epsilon^2} = \overline{x_d{}^2} - h_0 \overline{x_d} - \int h_1\, \overline{xx_d}\, d\alpha - \iint h_2\, \overline{xxx_d}\, d\alpha\, d\beta$$

which is the same as Eq. (46) and proves our claim.

ACKNOWLEDGMENT

It is a pleasure for the authors to acknowledge many interesting and helpful discussions that Mr. Katzenelson had with Professor N. Levinson of the Department of Mathematics, M.I.T. and Dr. H. Furstenberg, formerly with the Department of Mathematics, M.I.T. and now at the University of Minnesota.

REFERENCES

BAGDADI, E. J. (1961). "Lectures on Communication Systems Theory," Chapter III, by W. M. SIEBERT. McGraw-Hill, New York.

BARRETT, J. F. The use of functionals in the analysis of nonlinear physical systems. Statistical Advisory Unit Report No. 1/57, Ministry of Supply, Great Britain (no date given).

BARRETT, J. F. AND LAMPARD, D. G. (1955). An expansion for some second-order probability distributions and its application to noise problems. *IRE Trans. of Information Theory* **IT-1,** 10.

BODE, H. W., AND SHANNON, C. E. (1950). A simplified derivation of linear least mean square smoothing and prediction theory. *Proc. IRE* **38,** 417–425.

BOSE, A. G. (May 1956). A theory of nonlinear systems. Technical Report 309, Research Lab. of Electronics, M.I.T.

BRILLIANT, M. B. (March 1958). Theory of the analysis of nonlinear systems. Technical Report 345, Research Lab. of Electronics, M.I.T.

CHESLER, D. A. (January 1960). Nonlinear systems with gaussian inputs. Sc.D. Thesis, Dept. of Elect. Eng., M.I.T.

COURANT, R. AND HILBERT, D. (1953). "Methods of Mathematical Physics," Vol. I, Chapter III, pp. 112–163. Interscience, New York.

CUNNINGHAM, N. J. (1958). "Introduction to Nonlinear Analysis." McGraw-Hill, New York.

DAVENPORT, W. B., JR. AND ROOT, W. L. (1958). "An Introduction to the Theory of Random Signals and Noise," Chapter II, pp. 230–231. McGraw-Hill, New York.

DOOB, J. L. (1953). "Stochastic Processes." Wiley, New York.

FRECHET, M. (1910). Sur les Fonctionnelles Continues. *Ann. Ec. Norm. Sup.* **27,** 193–219.

GEORGE, D. A. (July 1959). Continuous nonlinear systems. Technical Report 355, Research Lab. of Electronics, M.I.T.

KOLMOGOROV, A. N. AND FOMIN, S. V. (1957). "Elements of the Theory of Functions and Functional Analysis," Vol. I, Addendum to Chapter III, p. 105. Gray Book Press, Rochester, N. Y. Translated from Russian by L. F. Born.

LEE, Y. W. (1960). "Statistical Theory of Communication." Wiley, New York.

LEVINSON, N. A Heuristic Exposition of Wiener's Mathematical Theory of Prediction and Filtering, Appendix C to Wiener (1949).

MINORSKY, N. (1957). "Introduction to Nonlinear Mechanics." Edwards, Ann Arbor, Mich.

NEWTON, G. C., GOULD, L. A. AND KAISER, J. F. (1957). "Analytical Design of Linear Feedback Controls." Wiley, New York.

NUTTALL, A. H. (May 1958). Theory and application of the separable class of random processes. Technical Report 343, Research Lab. of Electronics, M.I.T.

SINGLETON, H. E. (August 1950). Theory of nonlinear transducers. Technical Report 160, Research Lab. of Electronics, M.I.T.

SMITHIES, F. (1958). "Integral Equations," p. 164. Cambridge Univ. Press.

VOLTERRA, V. (1930). "Theory of Functionals and of Integral and Integro-Differential Equations." Blakie and Sons, Ltd. London. Dover, New York, 1959.

WIENER, N. (1930). Generalized harmonic analysis. *Acta Math.* **55,** 117–258.

WIENER, N. (April 1942). Response of a nonlinear device to noise. Report V-16S, Radiation Laboratory, M.I.T.

WIENER, N. (1949). "Extrapolation, Interpolation and Smoothing of Stationary Time Series," p. 74. Wiley, New York.

WIENER, N. (1958). "Nonlinear Problems in Random Theory." The Technology Press, M.I.T., and Wiley, New York.

ZAMES, G. (August 1960). Nonlinear operators for systems analysis. Sc.D. Thesis, Dept. of Elec. Eng., M.I.T.

ZAMES, G. (July 1961). On the mapping of random processes into random variables in the unit interval and their characterization by probability distribution operators. ARM-258, Applied Research Laboratory, Sylvania Electronic Products, Waltham, Mass.

ZADEH, L. A. (1953). Optimum nonlinear filtering. *J. Appl. Phys.* **24,** 396.

ZADEH, L. A. (1953). A contribution to the theory of nonlinear systems. *J. Franklin Inst.* **255,** 387–408.

ZADEH, L. A. (1957). On the representation of nonlinear operations. *IRE Wescon Conv. Record,* **Pt. 2,** 105.

ZADEH, L. (1961). "Progress in Information Theory in U.S.A. 1957–60," Chapter V: Prediction and filtering. *IRE Trans. on Information Theory,* **IT-7,** 128.